

Towards Transparency by Design for Artificial Intelligence

Year: 2020 | Citations: 342 | Authors: Heike Felzmann, E. Fosch-Villaronga, C. Lutz, Aurelia Tamó-Larrieux

Abstract

In this article, we develop the concept of Transparency by Design that serves as practical guidance in helping promote the beneficial functions of transparency while mitigating its challenges in automated-decision making (ADM) environments. With the rise of artificial intelligence (AI) and the ability of AI systems to make automated and self-learned decisions, a call for transparency of how such systems reach decisions has echoed within academic and policy circles. The term transparency, however, relates to multiple concepts, fulfills many functions, and holds different promises that struggle to be realized in concrete applications. Indeed, the complexity of transparency for ADM shows tension between transparency as a normative ideal and its translation to practical application. To address this tension, we first conduct a review of transparency, analyzing its challenges and limitations concerning automated decision-making practices. We then look at the lessons learned from the development of Privacy by Design, as a basis for developing the Transparency by Design principles. Finally, we propose a set of nine principles to cover relevant contextual, technical, informational, and stakeholder-sensitive considerations. Transparency by Design is a model that helps organizations design transparent AI systems, by integrating these principles in a step-by-step manner and as an ex-ante value, not as an afterthought.